CONFIRMATORY FACTOR ANALYSIS (CFA) DIAGNOSTIC TESTS OF MATHEMATICS LEARNING DIFFICULTIES OF ELEMENTARY STUDENT

[1]Muhamad Arfan Septiawan, [2]Samsul Hadi

[1]Program Graduate School of Yogyakarta State University
Smarfan020@gmail.com
[2]Department of Education and Evaluation Research, Yogyakarta State University
samsul.d@gmail.com

Abstract

Mathematics is a highly structured knowledge. Mathematical skills should be built from the previous skills, the prerequisite skills must be completed before moving on to the next material. If misconception occurs in the previous material it will cause difficulties in the next level. Several instruments on diagnostic tests of mathematics learning difficulties have been developed, but most existing instruments are subjective tests like essays form that have limitation of material tested, and also require considerable time in checking student's answers. However, Kuesari (2014) in his dissertation developed an objective test instrument refers to the scoring of politomus. The purpose of this research is to analyze the validity of construct of diagnostic test instrument of mathematics learning difficulties which have been developed. Instruments were distributed to 184 students in fifth elementary grade. Confirmatory Factor Analysis is used to see a good models on diagnostic instruments. This validity tries to know the suitability of the item with the indicator to measure diagnostic tests of mathematics learning difficulties of elementary students in Yogyakarta.

Keywords: Diagnostic, learning difficulties, politomus, Construct Validity, CFA

INTRODUCTION

The results of the survey of the Program for International Student Assessment (PISA) show that the ability of Indonesian children in the world in unsatisfactory ranking. As released in the Litbang Kemendikbud page, Indonesia always below the international average score of 500. In 2000, Indonesia ranked 39 out of 41 countries with an average score of 367. In 2003, Indonesia was ranked 38th out of 40 countries with an average of 361. In 2006, Indonesia was ranked 50th out of 57 countries with an average score of 391. In 2009, Indonesia was ranked 61 out of 65 countries with an average score of 371. In 2012, in math competency only increased 375points and in 2015 up to 386 points with rank 62 of 72 countries (Litbang Kemendikbud, 2016). Although it has show an enhancement but still far from the average international standard.

The results of the above achievements can not be separated from the students' mathematical abilities from elementary schools ranging from know concept, concept structure and seek the relationship between the concept and structure.

Therefore, a test instrument is needed to diagnose the difficulty of learning mathematics in elementary school in order to know the weakness and ability in the mastery of mathematics. Some researchers have developed a diagnostic instrument to see the strengths and weakness of mathematical mastery. Sugiharto (2003) used essay test with 25 items. Kusmanto (2003) used a short essay test followed by questionnaires and interviews. Sidauruk (2005) in searching for misconceptions using multiple choice test but equipped with reasons. Fauzan (2010) developed a diagnostic test of learning difficulties using multiple-choice tests. Waskito & Kumaidi developed a model of analysis to identify the learning difficulties of students by compiling objective test items form multiple choice and essay.

However, the form of multiple choice instruments and essays is not yet quite effective because it takes a long time to correct and can not cover all the material in a single test. Kuesari (2014) in his dissertation developed a diagnostic test of mathematical difficulties based on the scoring of politomus. The instrument is based on the polarization of the student's answer, then each option has a score according to its level according to the scoring of politomus.

In this paper, the purpose of this research is to analyze the validity of construct of diagnostic test instrument of learning difficulties of mathematics that have been developed. Confirmatori factor analysis is used to see good models on diagnostic instruments (Adediwura, A. A., 2011, Hendryadi& Suryani, 2014). This validity tries to determine the suitability of items with indicators to measure diagnostic tests of mathematics learning difficulties elementary students. There are four indicators to measure the difficulty of learning mathematics, namely mastery in aspects of concepts, procedures, counting and representation.

The rest of this paper is organized as follow. Section 2 presents a description of the proposed methodology. Section 3 presents the simulation results following by discussion. Finally the conclusion and future works are presented in Section 4.

RESEARCH METHOD

This research was conducted as a quantitative descriptive to explore the construct validity of the diagnostic test of mathematics learning difficulties for elementary school students in Yogyakarta. A total of 184 elementary students participated in this study. The respondents were elementary students at 5 grades in 7 public elementary schools in Sleman regency.

Instruments amounted to 40 items, but only 32 were used and eliminated eight items to make the instrument more manageable and prevent negative impact on completion by the respondent without sacrificing its validity.

The study was conduct at February 2018 in 7 public elementary school in Sleman Regency, Special Region of Yogyakarta.

Using cluster random sampling out of 14 public elementary schools in Sleman Regency as population, 7 elementary schools were chosen, and reaching 184 elementary students at 6 grades as respondent.

Procedure and Data Analysis

The diagnostic test instrument is 40 item questions, then reduced to 32 items and eliminates eight items without sacrificing its validity to prevent drop outs during data collection. The last instrument is then distributed to the respondent.

The data then analysed using Lisrel 8.80 to measure the construct validity (Hendryadi & Suryani, 2014). Confirmatory factor analysis was used to analyse the construct validity of the data. Firstly, the overall model fit was conduct to examine the fit of the model based on the goodness fit indices then measure the measurement of fit. The criteria that were used to evaluate the goodness of fit are: normed chi-square ($\chi^2 / d.f.$), RMSEA (Root Mean Square Error of Approximation), RMR (Root Mean-square Residual), GFI (Goodness-of-Fit Index), NFI (Normed Fit Index), Non-Normed Fit Index (NNFI), and CFI (Comparative Fit Index). Loading factor and *t*-Value are criteria that were used to analyze the measurement model fit. The standard of significance value for the validity based on Hair, Black, Babin, Ander-son, & Tatham (2010) were "factor loadings ± .3 to .4 are minimally acceptable".

RESULT AND DISCUSSION

Before examine the construct validity of the instrument, reliability was assessed to confirm the internal consistency of instruments items. The Cronbach's alpha statistics for the instrument was .741. These results confirmed the adequacy of the internal consistency of the instrument.

To examine the overall and measurement model fit, confirmatory factor analysis was used. The explanation of each criteria in overall model fit that mentioned before i.e. Normed Chi-Square is ratio between Chi-Square and degree of freedom. RMSEA is most informative indicator for model fit. RMR represent the residual mean by matching the covariance matrix of the data. GFI is a scale of precision of the model that resulting covariance matrix. NFI has a tendency to lowering the fit in small sample size. NNFI was used to fix the problem that caused by the complexity of the model.

Overall Model Fit

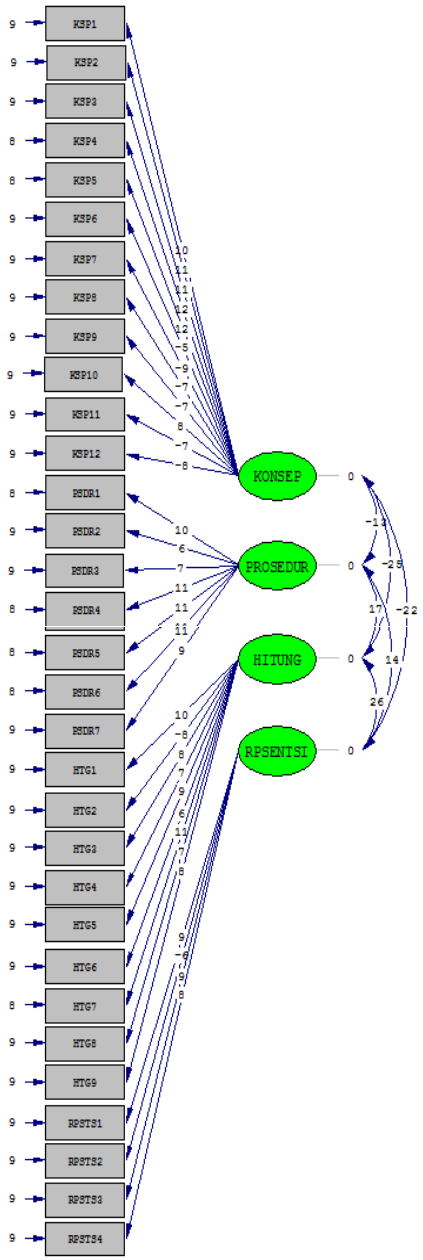
To analyse the construct validity, it necessary to the overall model of fit first. If overall model fit match with the criteria, then the measurement model fit could be conduct to fulfil the construct validity analysis condition.

Presented outputs and tables showing the fit values according to the overall model after the data were analyzed using Lisrel 8.80 (Hendryadi & Suryani, 2014).

No	GOF Criteria	Result	Level of Fit
1	Normed χ^2	3.11	Poor fit
2	RMSEA	.107	Poor fit
3	RMR	.089	Marginal fit
4	GFI	.67	Good fit
5	NFI	.88	Marginal fit
6	NNFI	.91	Marginal fit
7	CFI	.92	Marginal fit

Table 1 shows the normed $\chi 2$, that is the ratio between the $\chi 2$ and degree of freedom. Good fit level suggests that the score must be range from 1.00 to 2.00. because the score is 3.11 so that the normed $\chi 2$ is in the poor level. RMSEA scored .055. Because the score is greater than the suggested score (RMSEA< .05), so the level of fit is in poor fit (Browne & Cudeck, 1993). The result of RMR and CFI is .089 and .92 respectively, with N≤250 the model will in good fit level if Standardized RMR ≤ .09 and CFI> .92 (Hair, Black, Babin, Anderson, & Tatham, 2006) so that those criteria are in marginal fit. The criteria of GFI range from 0 (poor fit) to 1 (perfect fit). The GFI scores .67 so that the fit level

is in good criteria. The criteria of GFI is same for NFI and NNFI. Considering the condition above, it could say that the overall model is fit to measure the construct validity of the instrument.



(based in table above we can use GOF Criteria RMR, GFI, NFI, NNFI, CFI to assumed that the overall model is fit, so we can go to the next step to measurement the contruxt validity of the instrument that the level is marginal and good fit. (Why the marginal can be use to measure, because the marginal fit is closely fit) Marginal fit = fit at the boundary.)

Measurement Model Fit

Once the overall model is appropriate, a suitable measurement model is performed. The results of the appropriate measurement model for diagnostic tests are described here. Appropriate measurement models are performed using first-order or first CFA confirmatory factor analysis. Criterion of construct validity is loading factor and t-Value. If the loading factor is greater than 0.3 and t-Value over 1,96 then the item is categorized as a valid item. The criterion refers to the assumption that "a factor loading of ± 0.3 to .4 is minimally acceptable." (Hair, Black, Babin, Anderson, & Tatham, 2006).

Factor	Item	First Order CFA		Interpret
		LF	t-Val	
KONSEP	KSP1	.80	10.28	Valid
	KSP2	.79	10.91	Valid
	KSP3	.81	11.20	Valid
	KSP4	.76	11.82	Valid
	KSP5	.83	11.57	Valid
	KSP6	-.46	-5.33	Not Valid
	KSP7	-.57	-9.18	Not Valid
	KSP8	-.50	-6.85	Not Valid
	KSP9	-.49	-7.22	Not Valid
	KSP10	.61	7.87	Valid
	KSP11	-.49	-6.63	Not Valid
	KSP12	-.51	-7.98	Not Valid
PROSDUR	PSDR1	.52	9.82	Valid
	PSDR2	.30	5.84	Valid
	PSDR3	.39	7.04	Valid
	PSDR4	.68	10.53	Valid
	PSDR5	.66	10.83	Valid
	PSDR6	.68	10.67	Valid
	PSDR7	.59	8.70	Valid
HITUNG	HTG1	.69	10.40	Valid
	HTG2	-.73	-8.02	Not Valid
	HTG3	.58	7.74	Valid
	HTG4	.60	7.74	Valid
	HTG5	.71	9.04	Valid
	HTG6	.47	6.09	Valid
	HTG7	.74	10.97	Valid
	HTG8	.44	6.77	Valid
	HTG8	.57	8.43	Valid
RPESNTSI	RPSTS1	.62	9.37	Valid
	RPSTS2	-.52	-5.83	Not Valid
	RPSTS3	.65	8.99	Valid
	RPSTS4	.55	8.43	Valid

Table 2 shows a summary of the construct validity using the first CFA. This indicates that the instrument item has loading factor> .3 and t-Value> 1.96 means the instrument is valid constructively. Items KSP6, KSP7, KSP8, KSP9, KSP11, KSP12, HTG2, RPSTS2 are not constructively valid to

diagnose the difficulty of learning mathematics from elementary school students.

CONCLUSION AND SUGGESTION

Based on the analysis results, it can be concluded that the conclusion of 24 items has a loading factor> .3 and t-value> 1.96 so that of 32 items only 24 valid items. 2) 24 diagnostically adapted diagnostic items apply to measure students' ability to master mathematics from elementary school students in the Yogyakarta context. Diagnostic diagnostic test instruments are very useful, both for teachers, who want to monitor students' mathematical abilities, and for researchers, who often use different instruments in their studies.

REFERENCES

Adelson, J.L., & McCoach, D.B. (2010). Measuring the mathematical attitudes of elementary students: The effects of a 4-point or 5-point Likert-type scale. *Educational and Psychological Measurement no.70, vol* 5, 796-807.

Adediwura, A. A. (2011). The development and confirmatory factor analysis of a scale for the measurement of gifted students attitude towards mathematics. World Journal of Education, no.1, vol 1, 52–62.

Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J.S Long (Eds), *Testing Structural Equatin Models*. Newbury Park, CA: Sage, 136-162 (Electronic Version).

Fauzan, (2010). *Pengembangan tes diagnostik kesulitan belajar matematika di SMA*. Disertasi Doktor, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.

Hair, Jr.J.F., Black, W.C., Babin, B.J., Anderson, R.E. & Tatham, R.I. (2006). *Multivarite Data Analysis, 6th edition.* New Jersey: Person Prentice Hall.

Hair, J.F., Anderson, R.E., Tatha,, R.L. & Black, W.C. (2010). *Multivariate data analysis* (7th ed). New Jersey: Prentice-Hall.

Hendryadi & Suryani, 2014. *Structural Equation Modeling dengan Lisrel 8.80.* Yogyakarta: Kaukaba Dipantara.

Kuesari (2014). *Pengembangan tes diagnostik kesulitan belajar matematika di SD.* Disertasi Doktor, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.

Kusmanto, (2003). *Peningkatan efektifitas pembelajaran matematika di SLTP dengan diagnostik kesulitan belajar dan remidial melalui penelitian tindakan.* Tesis Magister, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.

Puspendik 2016. "survei PISA. Diambil pada tanggal 3 maret 2018, dari http://litbangkemendiknas.net/detail.php/id=214.

Sidaruk, S, (2005). *Miskonsepsi stoikimetri pada siswa*. Disertasi doktor, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.

Sugihartono, (2003). *Diagnostik kesulitan siswa SMU dalam menyelesaikan soal matematika.* Tesis Magister, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.

Waskito & Kumaidi, (2010). *Asesmen untuk memantau kualitas pendidikan pengembangan model analisis prestasi belajar untuk mengidentifikasi kesulitan belajar peserta didik.* Jakarta: UIN Syarif Hidyatullah, Jakarta.